Response Letter

Sincerest thanks for the comments on our manuscript entitled “Improving Cloud Type Classification of Ground-Based Images Using Region Covariance Descriptors” (amt-2020-189). These comments are very valuable and helpful for revising and improving our paper, and they also have important guiding significance to our research. We have studied comments carefully and made revisions which we hope meet with approval.

The main revisions in the paper and the responses to the reviewer’s comments are as following:

Reviewer #3:

The manuscript has structural issues, unclear formulations and lacks details in parts which hampers its ability to communicate the research to the reader. The manuscript lacks evidence of the physical relevance of the particular cloud types used - how specifically are these cloud types important for climate predictions and weather forecasting? In addition, there are structural issues in the manuscript, for example many sentences combine terms that are not on equal footing and often it is not detailed how and why specific assertions are evidenced. For example, in the introduction’s first sentence: it is true that "clouds have a strong impact" on "Earth’s energy budget", but it isn’t clear how clouds having a "strong impact" on "climate modelling" or "weather prediction". Further, line 24 goes on to talk about "weather monitoring" rather than "modelling" which are not equivalent. Finally, an extensive list of publications covering "cloud coverage measurement" and "cloud classification", but this publication doesn’t appear to be about "cloud coverage". The fact that there is "additional interest" doesn’t evidence that "cloud type classification" is in great need, or is that in fact what these papers state?

Reply: Many thanks for the comment and we have added the content of how cloud types affect climate change in Introduction Section of the revised manuscript. Introduction is modified as follows:

“Clouds affect the Earth’s climate by modulating Earth's basic radiation balance (Hartmann et al., 1992; Ramanathan et al., 1989). Cloud type variations are shown to be as important as cloud cover in modifying the radiation field of the earth–atmosphere system. For example, stratocumulus, altostratus, and cirrostratus clouds produce the largest annual mean changes of the global top-of-atmosphere and surface shortwave radiative fluxes (Chen et al., 2000). Cloud type is also one of the most reliable predictors of weather, e.g., cirrocumulus clouds are a sign of good weather. Therefore, accurate cloud type classification is in great need. Currently, the classification task is mainly undertaken by manual observation, which is labour-intensive and time-consuming. Benefiting from the development of ground-based cloud image devices, we are able to continuously acquire cloud images and automatically classify the cloud types.”

The performance analysis is a bit weak as the datasets used are quite small in size (784 for SWIMCAT *and 500 zenithal). I would suggest using data-augmentation (random rotations and zoom) to create two orders of magnitude more images to work with.*

Reply: Thanks for your concern. Data-augmentation is a great way to expand a limited dataset, especially for deep learning, however, it's not a magic bullet. One always runs the risk of overfitting the data model to the training samples if relying too much on data augmentation technique. If the dataset size expands to create two orders of magnitude more images, the distribution of the sample features may make the model deviate from the real data.

To validate the assumption, we conduct an experiment, in which we compare the predict accuracy of our model trained by data with/without augmentation. Let M1 and M2 denote the proposed model trained without and with data augmentation, respectively. In the training stage of M2, we randomly rotate and flip each training image to generate 9 more images, as shown in Fig. 1 and Fig. 2. The average accuracy obtained by 10-fold cross validation is reported in Table 1. The experiment result shows that, compared to M2, M1 acquires higher recognition accuracy on both datasets. The reason is that the RCovDs used in our model owns rotation and scale invariance characteristics. Moreover, the BoF method makes no use of the location information of RCovDs. Therefore, based on the experimental result, it seems that the data augmentation technique does not significantly improve the accuracy of the proposed model. Others, as mentioned in the paper, one purpose of our model is to address the dilemma that how we can obtain a high cloud type recognition accuracy even with a limited training set.

Table 1. The accuracy of M1 and M2 test on SWIMCAT and *zenithal* dataset.

	Accuracy of SWIMCAT dataset	Accuracy of *zenithal* dataset
Model trained without data augmentation (M1)	**98.4%**	**98.6%**
Model trained with data augmentation (M2)	97.7%	94%

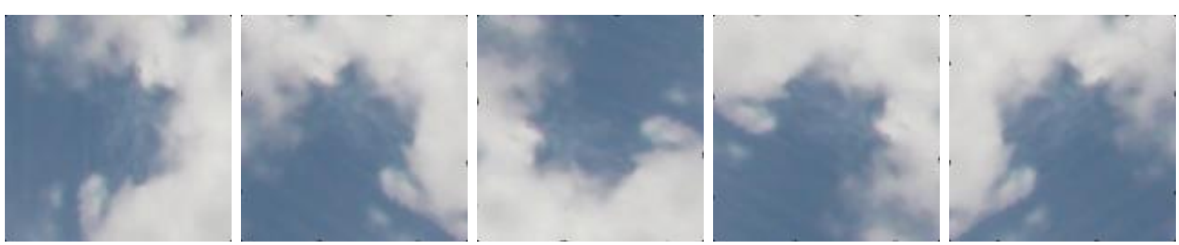


Figure 1 Samples of SWIMCAT data augmentation. The upper left image is the original image, and the other images are generated by random flip and rotation.

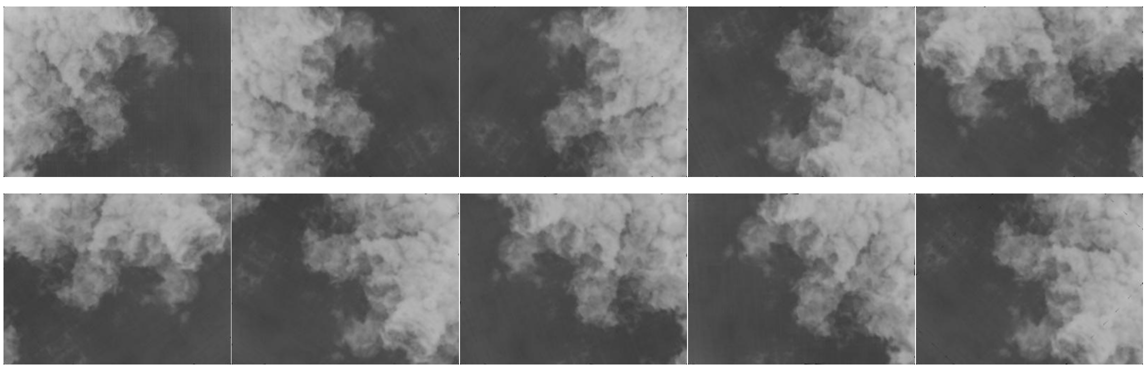


Figure 2 Sample of *zenithal* data augmentation. The upper left image is the original image, and the other images are generated by random flip and rotation.

In addition, because random subsets of the datasets are used for training and validation it is unclear to me whether the performance difference to prior work is actually due to random chance (training/testing on easier partition if the datasets) or whether the technique presented here is indeed better. It is common in machine learning datasets to train all models on the same training set and validate against the same validation set.

Reply：To evaluate the performance of machine learning model，we need to test it on unknown data (or first seen data). Cross validation is one of the techniques used to test the effectiveness of a machine learning model. To perform cross validation, we need to keep aside a sample/portion of the data on which is not used to train the model, later use this sample for testing/validating. The commonly used cross validation techniques are ***holdout method*** and ***k-fold cross validation*.**

In the holdout method, we randomly split complete data into training and test sets, ideally split the data into 70%:30% or 80%:20%. Then perform the model training on the training set and use the test set for validation purpose. In typical cross-validation, results of multiple runs of model-testing are averaged together; in contrast, the holdout method, in isolation, involves a single run. The disadvantage of the holdout method is that the performance evaluation is sensitive to the split ratio of the training set and the validation set. It should be used with caution because without such averaging of multiple runs, one may achieve highly misleading results. If our data is huge and our test samples and train samples have the same distribution, then this approach is acceptable. Therefore, this method is usually used in deep learning models with **large scale dataset**.

In *k*-fold cross validation is a popular technique that generally results in a less biased model. Because it ensures that every observation from the original dataset has the chance of appearing in training and test set. This is one among the best approach if we have a **limited input data**. This method follows the below steps: (1) Split the entire data randomly into *k* folds. (2) Then train the model using the *k*-1 folds and validate the model using the remaining *k*-th fold. (3) Repeat this process until every *k* -fold serve as the test set. Then take the average of the recorded scores. That will be the performance metric for the model.

In our experiment, as the data size is limited, we choose *k*-fold cross validation for model validating. All comparison methods are conducted under the exact same training/test set partition during each k-fold cross validation process. There won't be a convincing problem since we repeat 50 times k-fold cross validation and take the average accuracy as final result.

Specific comments:

Abstract:

l 10: "Cloud types are important indicators of ... short-term weather forecasting" – this sentence doesn't make sense. "Cloud types" can't "indicate" "weather forecasting. "The meteorological researchers can benefit from the automatic cloud type recognition of massive images captured by the ground-based imagers". Why is this true? Also, I would leave out "The" in "The meteorological" and the word "massive".

Reply: Thanks for your concern. Currently, the classification task is mainly undertaken by manual observation, which is labour-intensive and time-consuming. Benefiting from the development of ground-based cloud image devices, we are able to continuously acquire cloud images and *automatically* classify the cloud types. We will delete the first two sentences "Cloud types are important indicators of ... captured by the ground-based imagers.". Following sentences will be added at the beginning of the Abstract.

"The distribution and frequency of occurrence of different cloud types affects the energy balance of the earth. Automatic cloud type classification of images continuously observed by the ground-based imagers cloud help climate researchers uncover the relationship between cloud type variations and climate change."

l 12: "However, by far it is still of huge challenge to design a powerful discriminative classifier for cloud categorization" - why is it a huge challenge?

Reply: Thanks for your comment. The challenges are explained in Introduction Section. The main challenges of the ground-based cloud image classification task can be ascribed to the following reasons: (1) One single feature cannot effectively describe different types of clouds, we need to extract textural, structural, and statistical features simultaneously. (2) The scale of cloud varies greatly, therefore, the extracted features should be robust in the presence of illumination changes and nonrigid motion. (3) Different cloud types may have similar local characteristics, and thus the global features need to be considered.

l 14: "BoF is extended from Euclidean space to Riemannian manifold by k-Means clustering, in which Stein divergence is adopted as a similarity metric" - what is the relevance of this? Why is this done?

Reply: RCovDs are Symmetric Positive Defined (SPD) matrices and naturally reside in a Riemannian manifold, therefore, the traditional Euclidean distance metric is no longer applicable. In other words, in the process of k-Means clustering, when measuring similarity between two features, the traditional Euclidean distance metric will be replaced by the version of Riemannian manifold; that is Stein divergence in our paper. Also, the machine learning algorithms on Euclidean space should be adapted as well. Thus, we encode all RCovDs of one image into a histogram-like feature (vectorized feature) by using Riemannian counterpart of the conventional BoF, taking the geodesic distance of the underlying manifold as the metric.

l 15: "The histogram feature is extracted by encoding RCovDs of the cloud image blocks with BoF-based codebook" - the term "histogram feature" hasn't been explained yet, what is this? How is it relevant to the technique/results of this paper?

Reply: A bag of features is a vector of occurrence counts of a vocabulary of local image features, that is, a sparse histogram over the vocabulary. The specific steps of BoF method are as follows (Jégou et al., 2010):

1. Feature extraction: Extract local image descriptors of each image in the training set.
2. Quantization: Descriptors are quantized into visual words with the *k*-Means algorithm.
3. Image representation: An image is then represented by the frequency histogram of visual words obtained by assigning each descriptor of the image to the closest visual word.

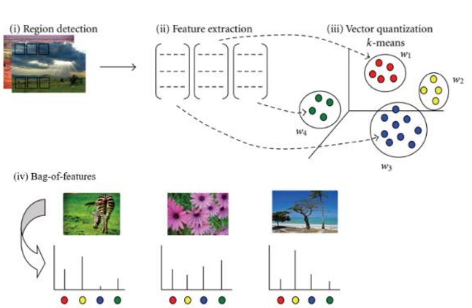


Figure 3 Four steps for constructing the bag-of-features for image representation.(Tsai, 2012)

The k-dimensional vector is called histogram feature. BoF describes an image as a vector from a set of local descriptors, and it aggregates the local features to obtain a global histogram representation. In fact, to encode and describe an image with the histogram vector, a good classification result can be achieved. Moreover, BoF has a dimensionality reduction effect, which will be much more effective for the subsequent SVM classification process.

l 17: "The experiments on the ground-based cloud image datasets validate the proposed method and exhibit the competitive performance against state-of-the-art methods." - this should instead specify exactly what the improvements on previous work is, give the numbers that indicate the improvement and what the implications of these improvements are.

Reply: Many thanks for the valuable comment. The last sentence of the Abstract will be changed as follows:

“The experiments on the ground-based cloud image datasets shows that a very high prediction accuracy (more than 98% on two datasets) can be obtained with a small number of training samples, which validate the proposed method and exhibit the competitive performance against state-of-the-art methods.”

General:

- the section on "Feature extraction" should be before "Region Covariance Descriptors" since the covariance descriptors used the features.

Reply: Thanks for your reminder. These two sections will be interchanged in the revised manuscript.

l 102: the relationship between w in the "Rectangular region R with size w x w" and the width of the input image isn’t specified.

Reply: Thanks for the comment. The relationship between *w* and rectangular region *R* is illustrated in **Section 3.1**.

Once again, thank you very much for your creative comments and suggestions.

References

Chen, T., Rossow, W. B., and Zhang, Y.: Radiative Effects of Cloud-Type Variations, J. Clim., 13, 264-286, 10.1175/1520-0442(2000)013<0264:reoctv>2.0.co;2, 2000.

Hartmann, D. L., Ockert-Bell, M. E., and Michelsen, M. L.: The Effect of Cloud Type on Earth's Energy Balance: Global Analysis, J. Clim., 5, 1281-1304, 10.1175/1520-0442(1992)005<1281:Teocto>2.0.Co;2, 1992.

Jégou, H., Douze, M., and Schmid, C.: Improving Bag-of-Features for Large Scale Image Search, Int. J. Comput. Vision, 87, 316-336, 10.1007/s11263-009-0285-2, 2010.

Ramanathan, V., Cess, R., Harrison, E. F., Minnis, P., Barkstrom, R. B., Ahmad, E., and Hartmann, D.: Cloud-radiative forcing and climate: Results from the Earth's radiation budget, Science, 243, 57-63, 10.1126/science.243.4887.57, 1989.

Tsai, C.-F.: Bag-of-Words Representation in Image Annotation: A Review, ISRN Artificial Intelligence, 2012, 376804, 10.5402/2012/376804, 2012.

Improving Cloud Type Classification of Ground-Based Images Using Region Covariance Descriptors

Yuzhu Tang[1, 3], Pinglv Yang[2], Zeming Zhou[2], Jianyu Chen[3], Delu Pan[3], Xiaofeng Zhao[2]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China
[2] College of Meteorology and Oceanology, National University of Defence Technology, Nanjing, 211101, China
[3] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Hangzhou, 310012, China

Correspondence to: Zeming Zhou (zhou_zeming@nudt.edu.cn), Jianyu Chen (chenjianyu@sio.org.cn)

Abstract. The distribution and frequency of occurrence of different cloud types affects the energy balance of the earth. Automatic cloud type classification of images continuously observed by the ground-based imagers cloud help climate researchers find the relationship between cloud type variations and climate change. However, by far it is still of huge challenge to design a powerful discriminative classifier for cloud categorization. To tackle this difficulty, in this paper, we present an improved method with region covariance descriptors (RCovDs) and Riemannian Bag-of-Feature (BoF). RCovDs model the correlations among different dimensional features, that allows for a more discriminative representation. BoF is extended from Euclidean space to Riemannian manifold by *k*-Means clustering, in which Stein divergence is adopted as a similarity metric. The histogram feature is extracted by encoding RCovDs of the cloud image blocks with BoF-based codebook. The multi-class support vector machine (SVM) is utilized for the recognition of cloud types. The experiments on the ground-based cloud image datasets shows that a very high prediction accuracy (more than 98% on two datasets) can be obtained with a small number of training samples, which validate the proposed method and exhibit the competitive performance against state-of-the-art methods.

1 Introduction

Clouds affect the Earth’s climate by modulating Earth's basic radiation balance(Hartmann et al., 1992; Ramanathan et al., 1989). Cloud type variations are shown to be as important as cloud cover in modifying the radiation field of the earth–atmosphere system. For example, stratocumulus, altostratus, and cirrostratus clouds produce the largest annual mean changes of the global top-of-atmosphere and surface shortwave radiative fluxes (Chen et al., 2000). Cloud type is also one of the most reliable predictors of weather, e.g., cirrocumulus clouds are a sign of good weather. Therefore, accurate cloud type classification is in great need. Currently, the classification task is mainly undertaken by manual observation, which is labour-intensive and time-consuming. Benefiting from the development of ground-based cloud image devices, we are able to continuously acquire cloud images and automatically classify the cloud types.

Clouds are by their very nature highly variable (Joubert, 1978), which makes the automatic classification a tough task. It is found that structure and texture are suitable to describe the visual appearance of clouds. The structural features include intensity gradient (Luo et al., 2018), mean grey value (Calbó and Sabburg, 2008; Liu et al., 2011), the census transform histogram (Xiao et al., 2016; Zhuo et al., 2014), edge sharpness (Liu et al., 2011), and features based on Fourier transform (Calbó and Sabburg, 2008). The textural features contain the scale invariant feature transform (SIFT) (Xiao et al., 2016), the grey level co-occurrence matrix (GLCM) (Cheng and Yu, 2015; Heinle et al., 2010; J. Huertas, 2017; Kazantzidis et al., 2012; Luo et al., 2018), the local binary patterns (LBP) (Cheng and Yu, 2015) and its extensions (Liu et al., 2015; Wang et al., 2018b). Commonly, no single feature is best suited for cloud type recognition, thus most existing algorithms tend to integrate multiple features to describe the cloud characteristics. However, those algorithms rarely consider the correlations between different dimensional features, which could lower the classification accuracy.

Within recent years, convolutional neural networks (CNNs) have been exploited to tons of image recognition and has achieved remarkable performance (Krizhevsky et al., 2012). Being different from hand-crafted features, CNNs extract hierarchical features including the low-level details and high-level semantic information. Recently, plenty of works (Shi et al., 2017; Ye et al., 2017) have obtained encouraging results by extracting the cloud signature from pre-trained CNNs, such as AlexNet (Krizhevsky et al., 2012) and VGGNet (Simonyan and Zisserman, 2015). In addition, attempts have been made to simply exploit end-to-end CNN models for cloud categorization (Li et al., 2020; Liu et al., 2019; Liu and Li, 2018; Liu et al., 2018; Zhang et al., 2018). However, the insufficiency of labelled samples might make the network hard to converge in the training stage.

The main challenges of the ground-based cloud image classification task can be ascribed to the following reasons: (1) One single feature cannot effectively describe different types of clouds, we need to extract textural, structural, and statistical features simultaneously. (2) The scale of cloud varies greatly, therefore, the extracted features should be robust in the presence of illumination changes and nonrigid motion. (3) Different cloud types may have similar local characteristics, and thus the global features need to be considered. To address those issues, we utilize the region covariance descriptors (RCovDs) to encode the features of the cloud image blocks, and with the aid of Bag-of-Feature (BoF), we aggregate those local descriptors to obtain the global cloud image feature for cloud type classification.

The performance of RCovDs (Tuzel et al., 2006) is proved to be superior on object detection (Carreira et al., 2015; Guo et al., 2010; Li et al., 2013; Pang et al., 2008) and classification tasks (Fang et al., 2018; Li et al., 2013; Wang et al., 2012). As the second-order statistics of the image features, RCovDs can provide rich and compact context representations. The noises are largely filtered out by removing the mean values of the features. RCovDs are also scale and rotation invariant, irrespective of the pixel positions and numbers of sample points. Despite of their attractive properties, directly adopting RCovDs for cloud type classification is still of difficulty on account of their non-Euclidean geometry property. RCovDs are Symmetric Positive Defined (SPD) matrices and naturally reside in a Riemannian manifold, therefore, the machine learning algorithms on Euclidean space should be adapted for the automatic cloud image recognition.

In Euclidean space, BoF describes an image as a vector from a set of local descriptors (Jégou et al., 2012), and it aggregates the local features to obtain a global representation. Inspired by the work in (Faraki et al., 2015a), we encode RCovDs of the local image blocks into a histogram by using Riemannian counterpart of the conventional BoF, taking the geodesic distance of the underlying manifold as the metric.

In this paper, we extend our previous work (Luo et al., 2018), and propose an improved cloud type classification method based on RCovDs. The diagram is shown in Fig. 1. In the first step, we extract multiple pixel-level features such as intensity, color and gradients from the cloud image blocks to form RCovDs. In the second step, RCovDs are encoded by the Riemannian BoF to output the histogram representation. In the last step, the histogram is taken as the feed of the multiclass SVM for cloud type prediction.

The main contributions of this paper are:

- The RCovD is firstly introduced to characterize the cloud image local patterns and the Riemannian BoF is applied to encode RCovDs into image-level histogram;
- The impacts of Riemannian BoF codebook size and the image block size on cloud type classification accuracy are investigated;
- For The small training dataset, the proposed algorithm offers better performance as compared to the state-of-the-art approaches.

The remainder of this paper is organized as follows. Section 2 introduces the ground-based cloud image datasets and details the proposed cloud type classification method. Experimental results and comparisons with other methods are presented in Section 3. Section 4 concludes our contributions and discusses the future work.

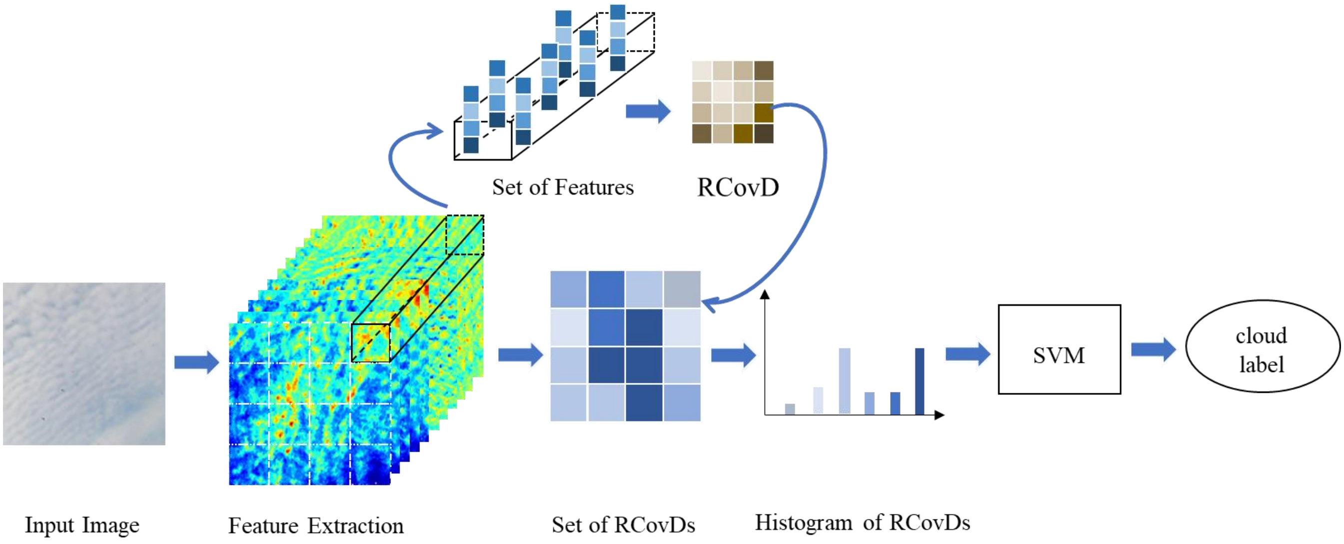


Figure 1: Pipeline of the proposed cloud classification method. Multiple pixel-level features are firstly extracted from the cloud image blocks to form RCovDs, then the histogram representation of RCovDs are obtained by Riemannian BoF, finally, the cloud type is predicted by multiclass SVM.

2 Data and methodology

2.1 Dataset

(1) SWIMCAT dataset: The Singapore Whole sky Imaging CATegories Database (SWIMCAT) was captured by Wide-Angle

High-Resolution Sky Imaging System (WAHRISIS) (Dev et al., 2014), a calibrated ground-based whole sky imager. During this observation, from January 2013 to May 2014, different weather conditions spanning several seasons are covered and a far-going cloud categories are collected. The SWIMCAT dataset involves 784 sky/cloud images, including 5 distinct classes: clear sky, patterned clouds, thick-dark clouds, thick-white clouds, and veil clouds. Figure 2 shows sample images from each category, the images have a dimension of 125×125 pixels (Dev et al., 2015).

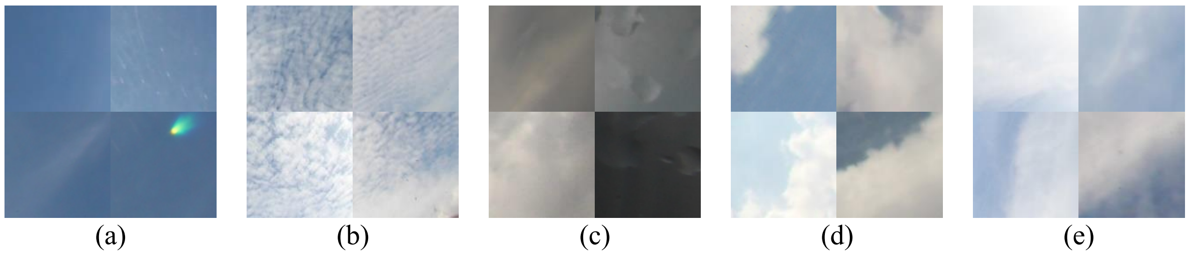


Figure 2: Sample images from the SWIMCAT dataset. The dataset includes five cloud types, namely, (a) Clear sky, (b) Patterned clouds, (c) Thick-dark clouds, (d) Thick-white clouds, and (e) Veil clouds.

(2) *zenithal* dataset: This dataset was acquired by the whole-sky infrared cloud-measuring system (WSIRCMS), which is located in Nanjing, China. The *zenithal* dataset contains 500 sky/cloud images, comprising of five different categories: cirriform clouds, clear skies, cumuliform clouds, stratiform clouds and waveform clouds (Liu et al., 2011; Liu et al., 2013). Figure 3 illustrates some sample images of different cloud types, and the image size is 320×240 pixels.

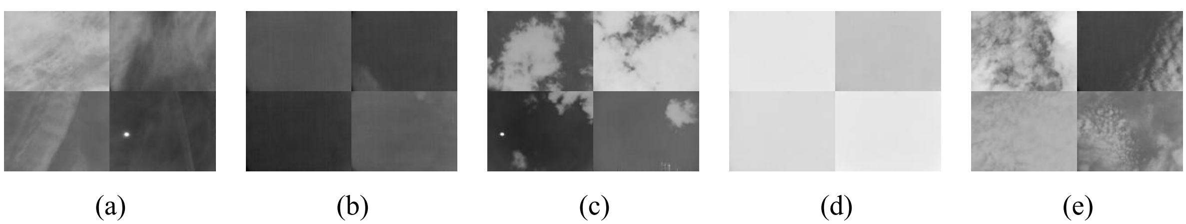


Figure 3: Sample images from *zenithal* dataset. (a) Cirriform clouds, (b) Clear sky, (c) Cumuliform clouds, (d) Stratiform clouds and (e) Waveform clouds.

2.2 Feature Extraction

The features for cloud type recognition should be representative and discriminative. In this paper, for the *zenithal* dataset, 7 features are extracted, including the image intensity $I(x,y)$, the norms of first and second order derivatives of $I(x,y)$ in both x and y direction, and the norm of gradient. The *zenithal* cloud image is mapped to a 7-dimensional feature space:

$$f_z = \left[I \quad |I_x| \quad |I_y| \quad |I_{xx}| \quad |I_{xy}| \quad |I_{yy}| \quad \sqrt{|I_x|^2 + |I_y|^2} \right]^T \tag{1}$$

As for the SWIMCAT dataset, we empirically choose the grayscale of B component, norms of first order derivatives of each color component, and the norm of gradient. Each pixel of the SWIMCAT image is transformed to a 13-dimensional feature map.

$$f_s = [B \quad |R_x| \quad |R_y| \quad |R_z| \quad |G_x| \quad |G_y| \quad |G_z| \quad |B_x| \quad |B_y| \quad |B_z| \quad \sqrt{|R_x|^2+|R_y|^2+|R_z|^2} \quad \sqrt{|G_x|^2+|G_y|^2+|G_z|^2} \quad \sqrt{|G_x|^2+|G_y|^2+|G_z|^2}\,]^T \tag{2}$$

We divide the cloud image into image blocks and then compute the SPD matrices with the feature maps defined in Eq. (1) and Eq. (2). With the Riemannian BoF, those local feature descriptors in the form of SPD matrices are converted into a histogram feature vector, which is used for cloud type classification.

2.3 Region Covariance Descriptors

Let f be the $W \times H \times d$ feature map extracted from the cloud image I. For a given rectangular region R with size $w \times w$, it contains $n = w \times w$ pixels of d-dimensional feature vectors $\{f_i, i = 1, 2 \cdots n\}$. The RCovD is defined by a $d \times d$ symmetric covariance matrix $\mathbf{C}_R$:

$$\mathbf{C}_R = \frac{1}{n-1}\sum_{i=1}^{n}(f_i - \mu)(f_i - \mu)^T \tag{3}$$

where $\mu = \frac{1}{n}\sum_{i=1}^{n} f_i$ is the mean of the feature vectors.

The RCovD correlates different components of the feature vectors, the diagonal entry $\mathbf{C}_R(i,i)$ represents the variance of i-th components of n feature vectors, and the element $\mathbf{C}_R(i,j)$ denotes the covariance of i-th and j-th components. Specifically, the RCovD subtracts the mean of the feature vectors, so it can filter out the noise to a certain extent. Note that, there might be a slight chance that $\mathbf{C}_R$ is not strictly positive definite, in this particular case, the $\mathbf{C}_R$ could be converted into a symmetric positive definite (SPD) matrix by adding a regularization term λE, where λ is a tiny coefficient which is set to $10^{-4} \times trace(\mathbf{C}_R)$ and E is the identity matrix (Huang et al., 2018; Wang et al., 2012; Wang et al., 2018a).

RCovDs belong to SPD manifold, when it is endowed with a Riemannian metric, it forms a Riemannian manifold. Based on the metric, the geodesic distance can be induced to measure the similarity of the image features. The geodesic distance is the length of the shortest curve between two SPD matrices on SPD Riemannian manifold. The most common distance is the Affine Invariant Riemannian Metric (AIRM) (Pennec et al., 2006):

$$\delta_G(\mathbf{X}, \mathbf{Y}) = \left\| \log(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2}) \right\|_F \tag{4}$$

where $\|\bullet\|_F$ is the Frobenius matrix norm and $\log(\bullet)$ denotes the matrix logarithm. The matrix logarithm can be calculated by singular-value decomposition (SVD), let $\mathbf{A} = \mathbf{U}\sum\mathbf{U}^{\mathbf{T}}$ be the eigenvalue decomposition of a symmetric matrix, the logarithm of $\mathbf{A}$ is given by

$$\log(\mathbf{A}) = \mathbf{U}\log(\sum)\mathbf{U}^{\mathbf{T}} \tag{5}$$

However, AIRM is computationally demanding. Driven by such computational concerns, in this paper, we adopt the Stein divergence (Sra, 2012) as a Riemannian distance metric, which is defined as

$$\delta_S(\mathbf{X},\mathbf{Y}) = (\log\left|\frac{\mathbf{X}+\mathbf{Y}}{2}\right| - \frac{1}{2}\log|\mathbf{XY}|)^{\frac{1}{2}} \tag{6}$$

where $|\bullet|$ denotes det operator.

2.4 Riemannian Bag-of-Feature

BoF requires a codebook with k codewords, which are usually obtained by clustering local descriptors. To extend the conventional BoF from Euclidean space into SPD Riemannian manifold $\mathcal{M}$, two issues should be considered: (1) Construct a codebook $\mathbb{C} = \{\mathbf{C}_j\}_{j=1}^{k}$ from a set of training RCovDs $\mathbb{X} = \{\mathbf{X}_i\}_{i=1}^{M}$. (2) Obtain a k-dimensional histogram from a set of RCovDs $\mathbb{E} = \{\mathbf{E}_i\}_{i=1}^{N}$ with the codebook $\mathbb{C}$.

An alternative way to learn a codebook is to apply the conventional k-means on vectorized RCovDs in the tangent space (Faraki et al., 2015b), however, it neglects the non-Euclidean geometric structure of SPD matrices. Taking the Riemannian geometry of SPD matrices into consideration, a possible way is to compute the cluster centers with Karcher mean (Pennec, 2006). The Karcher mean finds a point that minimizes the following object function

$$\mathbf{C}_j^* = \arg\min_{\mathbf{C}_j} \sum_i {\delta_S}^2\left(\mathbf{X}_i, \mathbf{C}_j\right) \tag{7}$$

where δ_S is Stein divergence to measure the geodesic distance of $\mathbf{X}_i$ and the clustering center $\mathbf{C}_j$. Given the training set $\mathbb{X}$, the codebook $\mathbb{C}$ is initialized by randomly selecting k RCovDs from $\mathbb{X}$, and iteratively update the cluster centers using Eq. (7) until the average distance of each point $\mathbf{X}_i$ to its nearest cluster is minimized. The procedure is summarized in Algorithm 1. We choose the number of codewords empirically by considering the trade-off between classification accuracy and computation consumption, which will be detailed in Section 3.

Algorithm 1: ***k*-Means clustering for codebook learning**

Input:

- training set $\mathbb{X} = \{\mathbf{X}_i\}_{i=1}^{M}$, $\mathbf{X}_i \in \mathcal{M}$
- k, the number of clusters
- *nIter*, the maximum number of iterations

Output:

- codebook $\mathbb{C} = \{\mathbf{C}_j\}_{j=1}^{k}$, $\mathbf{C}_j \in \mathcal{M}$

1: Initialize the codebook $\mathbb{C} = \{\mathbf{C}_j\}_{j=1}^{k}$ by selecting k samples from $\mathbb{X}$ randomly.

2: **for** $t = 1 \rightarrow nIter$ do

Assign each point $\mathbf{X}_i$ to its nearest cluster in $\mathbb{C}$.

Recompute each cluster center $\mathbf{C}_j^*$ using Karcher mean by minimizing Eq. (7).

Compute the geodesic distance ε between new cluster center $\mathbf{C}_j^*$ and original cluster center $\mathbf{C}_j$.

If ε is less than a predefined threshold or t reaches the maximum number of iterations, then break the loop.

end for

After obtaining the codebook $\mathbb{C}$, the image-level feature can be expressed with the histogram H of RCovDs. In the most straightforward case, H can be yielded by hard assign each RCovD $\mathbf{E}_i$ to the closest codeword in $\mathbb{C}$ with Stein divergence. The *j*-th ($1 \le j \le k$) dimension of H denotes the number of RCovDs assigned to the *j*-th codeword. To demonstrate the significance of histogram feature generated by Riemannian BoF, we randomly select half of images in the SWIMCAT dataset and partition each 125×125 image into 25 non-overlapping image blocks of size 25×25 to extract the second-order tensor features in the form of RCovD. Then, we learn a codebook of 10 codewords with Algorithm 1. In the same way, we select 20 images of each cloud type from the remaining images in the SWIMCAT dataset to construct a set of RCovDs for test, and assign each RCovD to the nearest codeword to obtain the RCovD histogram of each cloud type. As shown in Fig. 4, RCovDs from different cloud types have obviously separable codeword distributions. RCovD distributions of clear sky, pattern and thick-dark clouds are relatively concentrated, while the distributions of thick-white and veil clouds are slightly scattered. In particular, the RCovDs of veil clouds and clear sky are assigned to almost the same codewords, which makes the categorization of these two types challenging. Overall, our proposed Riemannian BoF provides vectorized discriminative representation for the cloud classification task.

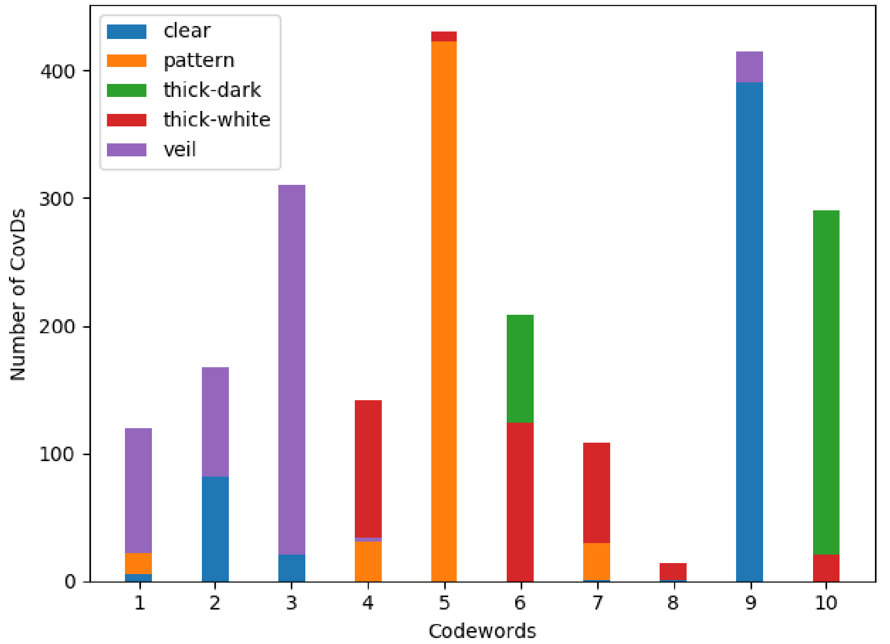


Figure 4: Histogram of RCovDs from different cloud types on SWIMCAT dataset. RCovDs from different cloud types have distinctive codeword distributions. RCovDs distributions of clear sky, pattern and thick-dark clouds are relatively concentrated, while the distributions of thick-white and veil clouds are slightly scattered. RCovDs of veils clouds and clear sky are assigned to almost the same codewords, which makes the categorization of these two types challenging.

2.5 Classification

SVM has significant performance in the classification task, since it establishes an input-output relationship straightly from the training dataset, and it exclude the need of any priori assumptions or specific preprocessing phases. Another merit is that, once the training procedure is finished, the classification is directly obtained in real time with a strong reduction of computation (Taravat et al., 2015).

For m-class classification tasks, there are several ways to build SVM classifiers. In this paper, the "one-against-one" method is adopted, in which $m(m-1)/2$ binary classifiers are constructed, and each classifier distinguishes one cloud type to another. We use the voting strategy to designate the cloud image to the category with the maximum number of votes (Chang and Lin, 2007; Hsu and Lin, 2002; Knerr et al., 1990; Kreßel, 1999). The proposed algorithm is summarized in Algorithm 2, in which SVM is implemented by the LIBSVM toolbox (Chang and Lin, 2007).

Algorithm 2: **The proposed cloud classification algorithm**

Input:

- Cloud image I with size $W \times H$
- Codebook $\mathbb{C} = \{\mathbf{C}_j\}_{j=1}^{k}$, $\mathbf{C}_j \in \mathcal{M}$

Output:

- Cloud type label L

Extract $W \times H \times d$ feature map f from cloud image I

Divide f into N blocks with size $w \times w \times d$, and construct RCovDs $\mathbb{E} = \{\mathbf{E}_i\}_{i=1}^{N}$ using Eq. (1)

Obtain a *k*-dimensional histogram $H = \{h_j\}_{j=1}^{k}$ representation of $\mathbb{E} = \{\mathbf{E}_i\}_{i=1}^{N}$:

  Initialize $H = \{h_j\}_{j=1}^{k}$ with zeros

  for $i = 1 \rightarrow N$ do

    Assign $\mathbf{E}_i$ to its nearest codeword $\mathbf{C}_j$, $h_j = h_j + 1$

  end for

Classify H using voting strategy:

  Initialize the number of votes $\{V_j\}_{j=0}^{m}$ with zeros

  for $i = 1 \rightarrow m(m-1)/2$ do

    Use the *i*-th binary SVM to classify H and obtain the prediction label L_j

    Update the number of votes by $V_j = V_j + 1$

  end for

Find the maximum number of votes V_j and output the corresponding label L_j

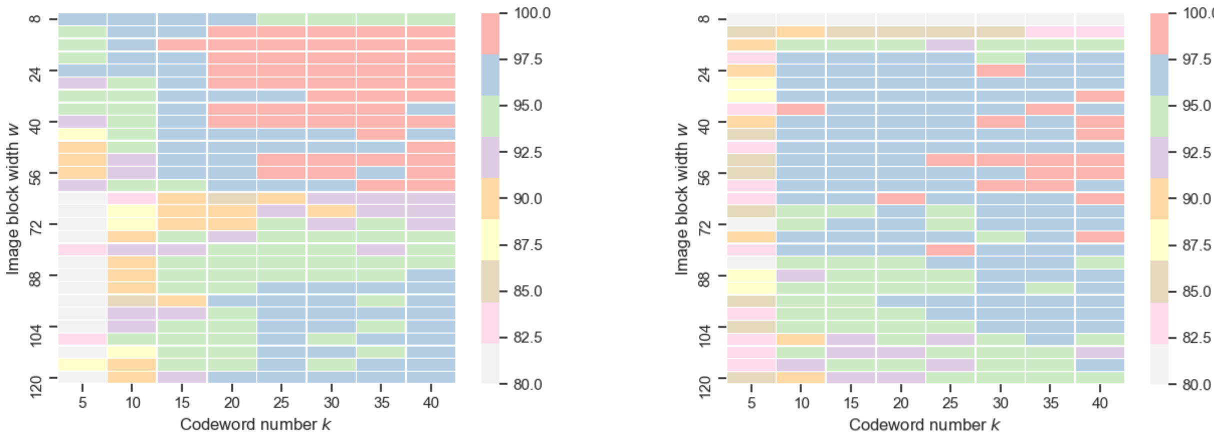
---

3 Experiments and discussion

To demonstrate the performance of our proposed cloud type classification method, we conduct several experiments on the SWIMCAT and *zenithal* datasets. We firstly analyze the effects of the two parameters (i.e. the codebook size *k* and the image block size $w \times w$) involved in the proposed algorithm on cloud type classification accuracy. Then, we design an empirical validation with various training/test partitions. Finally, we quantitatively evaluate and compare the best results of different methods, i.e., WLBP (Liu et al., 2015), BC (Cheng and Yu, 2015), and Luo's methods (Luo et al., 2018).

3.1 Parameter Configuration Analysis

In order to assess the impacts of the codebook size, i.e., the centroids number *k*, and the image block size $w \times w$ on cloud classification accuracy, we conduct sensitivity analysis on the SWIMCAT and *zenithal* datasets. In our experiments, *k* ranges from 5 to 40 with interval 5 and *w* ranges from 8 to 120 with the step size of 4. For a given *w*, the $W \times H \times d$ feature map is divided into $\lfloor \frac{W}{w} \rfloor \times \lfloor \frac{H}{w} \rfloor \times d$ blocks start from the upper left corner of the feature map and the incomplete blocks at the edges are dropped. We randomly choose 9/10 images of the dataset for training and the rest is for testing. The classification accuracy of each parameter configuration, as shown in Fig. 5, indicates that, to a certain extent, the larger the number of codebook size, the better performance on both datasets. However, we observe that the improvement is not statistical significance after *k* exceeds 20, while the computing burden increases obviously. In fact, the complexity of the Riemannian BoF is mainly determined by the cluster center number. We note that as the block size *w* increases, the classification accuracy increases first and then degrades beyond the highest point, this trend is especially evident on *zenithal* dataset. The reason is that larger blocks can capture more abundant texture information, while the local details might be ignored. Therefore, in the following experiments, considering trade-offs between classification accuracy and efficiency, we set $k = 30$, $w = 24$ for the SWIMCAT dataset, and *k* and *w* are set to 35 and 52 for the *zenithal* dataset.



(a) Sensitivity analysis on the SWIMCAT dataset (b) Sensitivity analysis on the *zenithal* dataset

Figure 5: Sensitivity analysis of parameter *k* and *w* in the proposed method on the SWIMCAT dataset and *zenithal* dataset.

3.2 Evaluation on Dataset with Small Sample Size

In machine learning tasks, suitable annotated data samples are in short supply and quite costly for classifier training and testing. Since manual labeling requires much workforce, it is of great significance to reduce the dependence of the classification model on the labeled dataset. To estimate the performance of the proposed method comprehensively, we extract different proportions of training images randomly from each dataset and take the rest images as the test set. In order to guarantee the stability of the classification results, each experiment was repeated five times to take the average as the final classification result. Figure 6 shows that in the situation of small sample size, for the SWIMCAT dataset, the proposed method achieves accuracy more than 90% on the test set with only 3% images (i.e., 24/784) of the dataset as the training set. The accuracy can be improved by 5% at least when the training set accounts for 9% images (i.e., 72/784). As for the *zenithal* dataset, our method obtains more than 90% classification accuracy on the test set when we randomly select 6% images (i.e., 30/500) of the dataset as training set, and achieving more than 95% accuracy when the proportion of training images increases to 10%. Generally, our proposed method significantly fulfills a high classification accuracy in small training sample situations. This is remarkable, considering that our proposed method is combining just RCovDs and Riemannian BoF. In conclusion, the proposed method requires only a few manually labeled samples to achieve a high cloud type recognition accuracy.

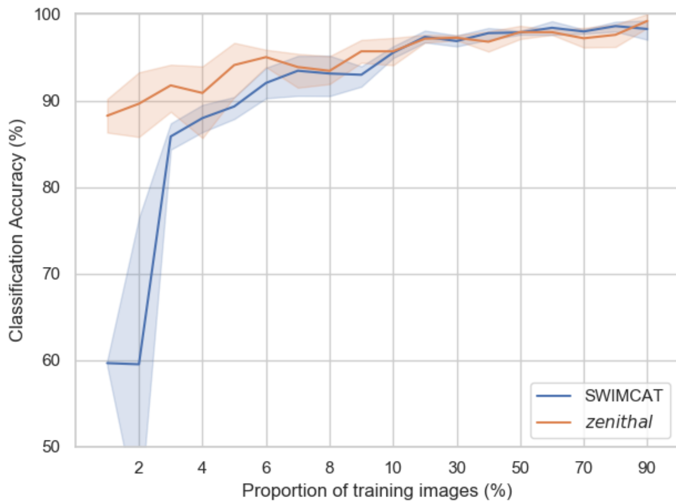


Figure 6: Performance analysis of training images with different proportions on the SWIMCAT dataset and *zenithal* dataset.

3.3 Comparison with state-of-the-art methods

Iterated cross validation is chosen as an effective scheme to verify the performance of the classifier. This strategy estimates the performance by randomly choose a part of the samples for independent training and testing the model without these samples, and repeating the procedure dozens of times (Beleites et al., 2013). In each experiment, we randomly select the same proportion (i.e., 1/10, 5/10, 9/10) of images for each category as the training set, and the remaining images are used as the test set. Each classification experiment is repeated 50 times to obtain the average accuracy as the final experimental result.

We compare the performance of our method with the best results published on the SWIMCAT dataset in Table 1.Notice that our algorithm utilizing RCovDs has a 2.58% accuracy rate at SWIMCAT dataset than other methods when the training sample accounts for 1/10 of the total data. And when the training sample accounts for 5/10 and 9/10, the proposed method is slightly higher than Luo's method and much higher than the other two methods. Figure 7 shows the confusion matrix of classification results with our proposed method on SWIMCAT dataset, with 9/10 of the dataset as training set. The discrimination rates of clear sky, pattern clouds and thick-dark clouds are perfect 100%, which demonstrates that these three types tend to be easily distinguished among all cloud types since they have the most significant features. Figure 8 shows two misclassified examples of SWIMCAT dataset, where yellow labels are the ground truth, and the red labels are the predicted cloud types predicted by our method. Notice that the veil clouds are prone to be misclassified as clear sky, since the veil clouds are thin and have highlight transmittance. Moreover, some veil clouds are misclassified as thick-white cloud, when the camera lens is contaminated, and the clouds is too thick. Besides, a small amount of thick-white clouds is misclassified as clear sky, pattern clouds or veil clouds.

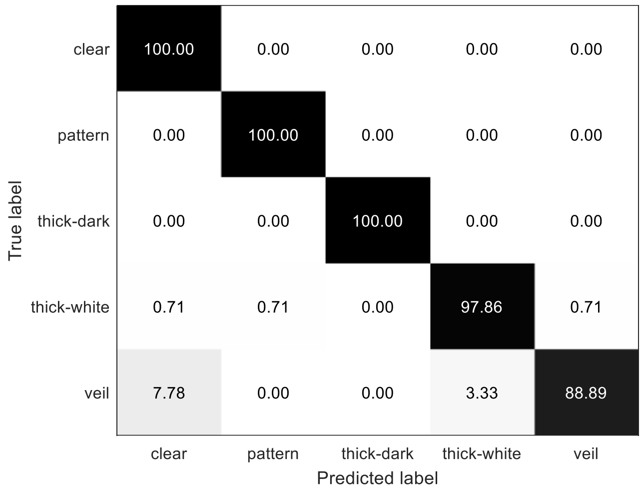


Figure 7: The confusion matrix of the SWIMCAT dataset classification results using our proposed method. 9/10 of the dataset is used for training and the rest is used for testing, the overall classification accuracy is 98.4%.

Figure 8: Misclassified images of SWMINCAT dataset. The yellow labels are the ground truth, and red labels are predicted cloud types. The veil clouds are prone to be misclassified as clear sky, since the veil clouds are thin and have high light transmittance, some veil clouds are misclassified as thick-white cloud, when the camera lens is contaminated and the clouds is too thick. Besides, a small amount of thick-white clouds is misclassified as clear sky, pattern clouds or veil clouds.

As for the *zenithal* dataset, Table 2 illustrates that the proposed method gains the highest overall accuracy compared with the other approaches. Figure 9 displays the confusion matrix of classification results with our method on the *zenithal* dataset, when 90% of the dataset is used as the training set. The discrimination rates of clear sky, cumuliform clouds and stratiform clouds are up to 100%. Only a small part of waveform clouds is misclassified as clear sky or cirriform clouds. In addition, some of the cirriform clouds are misclassified as clear sky or waveform clouds. Figure 10 illustrates the misclassified images of the *zenithal* dataset, waveform clouds and cirriform clouds are easy to be categorized as clear sky if the size of sky area is much larger than that of clouds. The reason why the waveform clouds and cirriform clouds are confused with each other is that they sometimes own extremely similar textures.

Table 1: Classification accuracy (%) of the SWIMCAT dataset obtained by different methods.

Method	1/10	5/10	9/10
WLBP**(Liu et al., 2015)**	72.31	84.52	88.86
BC**(Cheng and Yu, 2015)**	93.86	94.87	95.04
LUO**(Luo et al., 2018)**	91.83	97.72	97.86
Our method	**96.44**	**98.40**	**98.40**

Table 2: Classification accuracy (%) of the *zenithal* dataset obtained by different methods.

Method	1/10	5/10	9/10
WLBP**(Liu et al., 2015)**	81.64	92.24	93.48
BC**(Cheng and Yu, 2015)**	81.30	81.32	81.32
LUO**(Luo et al., 2018)**	90.85	95.98	96.36
Our method	**95.00**	**97.40**	**98.60**

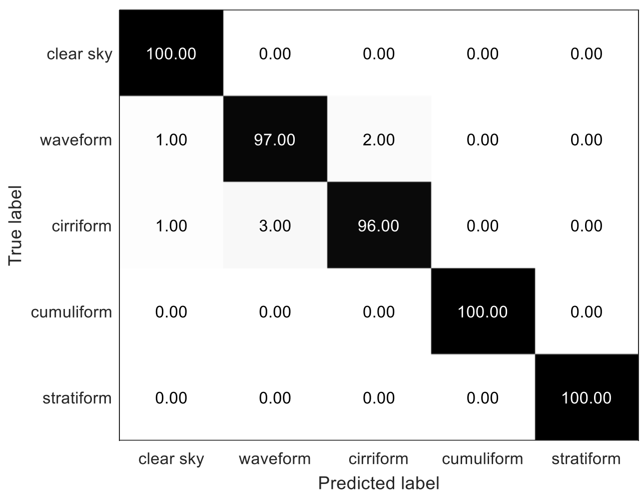


Figure 9: The confusion matrix of the *zenithal* dataset classification results using the proposed method. 9/10 of the dataset is used for training and the rest is used for testing, the overall accuracy is 98.6%.

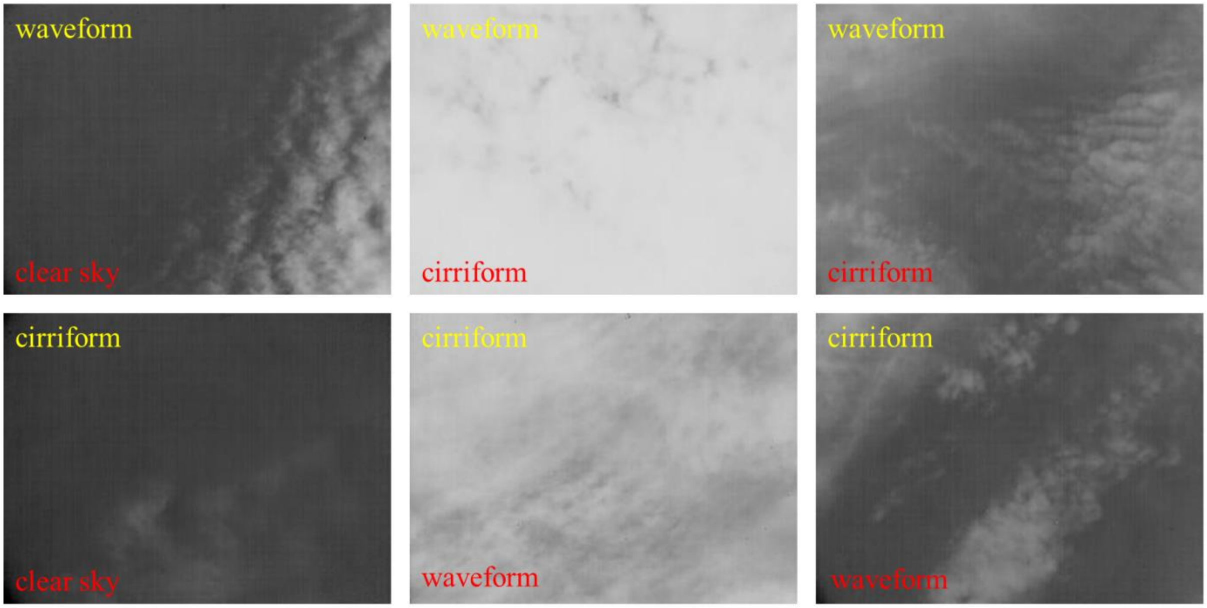


Figure 10: Misclassified images of the *zenithal* dataset. The yellow labels are the ground truth, and red labels are predicted cloud types. Waveform clouds and cirriform clouds are categorized as clear sky because the size of sky area is much larger than that of clouds, and these two cloud types are easily confused as they share similar local patterns.

4 Conclusions

To tackle the challenge of automatic cloud type classification for ground-based cloud images, in this paper, we present a new classification method with RCovDs as the local feature representation. RCovDs provide a simple way to fuse multiple pixel-level features, which improves discriminative ability for cloud images. The image-level information is obtained by applying Riemannian BoF to encode RCovDs into a histogram. Finally, we apply the “one-against-one” multi-class SVM as the classifier.

It is noted that even we choose relatively simple image features to calculate RCovDs, the performance of the proposed method is still impressive. We conduct parameter analysis experiment and figure out how block size *w* and codewords number *k* affect the accuracy of the proposed method. Classification experiments with different training set sizes demonstrate that our method is still efficient in the case of small size training set, which can greatly reduce the labor for labeling. In the third experiment, we compare our method to the other three cloud classification algorithms with different configurations of training/test sets. As the experimental results validate, the proposed method is competitive to state-of-the-art methods on both SWIMCAT and *zenithal* datasets.

In future work, the features like LBP or GLCM could be gathered and mapped into Riemannian manifold and multi-scale block strategy can be taken into consideration for a higher cloud type categorization accuracy. Others, the complex sky conditions with various cloud types should be deeply investigated to fulfill the application needs.

Code and data availability. The code of the proposed method can be made available via email to tangyuzhu9293@163.com. The SWIMCAT dataset used in this paper is available for download from http://vintage.winklerbros.net/swimcat.html, and the *zenithal* dataset can be made available via email to tangyuzhu9293@163.com.

Author contributions. YT performed the experiments and wrote the paper. PY analysed the data and designed the experiments. ZZ conceived the method and reviewed the paper. JC, DP and XZ reviewed the paper and gave constructive suggestions.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is supported by the National Key Research and Development Program of China (Grant No. 2016YFC1400903), NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (Grant No. U1609202) and the National Natural Science Foundation of China (Grant Nos. 61473310, 41174164, 41775027, 41376184 and 40976109).

References

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., and Popp, J.: Sample size planning for classification models, Analytica Chimica Acta, 760C, 25-33, 10.1016/j.aca.2012.11.007, 2013.

Calbó, J. and Sabburg, J.: Feature Extraction from Whole-Sky Ground-Based Images for Cloud-Type Recognition, J. Atmos. Oceanic Technol., 25, 3-14, 10.1175/2007JTECHA959.1, 2008.

Carreira, J., Caseiro, R., Batista, J., and Sminchisescu, C.: Free-Form Region Description with Second-Order Pooling, PAMI, 37, 1177-1189, 10.1109/TPAMI.2014.2361137, 2015.

Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol., 2, 1-39, 10.1145/1961189.1961199, 2007.

Chen, T., Rossow, W. B., and Zhang, Y.: Radiative Effects of Cloud-Type Variations, J. Clim., 13, 264-286, 10.1175/1520-0442(2000)013<0264:reoctv>2.0.co;2, 2000.

Cheng, H. Y. and Yu, C. C.: Block based cloud classification with statistical features and distribution of local texture features, Atmos. Meas. Tech., 8, 1173-1182, 10.5194/amt-8-1173-2015, 2015.

Dev, S., Lee, Y. H., and Winkler, S.: Categorization of cloud image patches using an improved texton-based approach, in: Proceedings of 2015 IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27-30 Sept 2015, 422-426, 2015.

Dev, S., Savoy, F. M., Lee, Y. H., and Winkler, S.: WAHRSIS: A low-cost high-resolution whole sky imager with near-infrared capabilities, in: Proceedings of SPIE - The International Society for Optical Engineering 9071:90711L on SPIE Defense + Security, Baltimore, Maryland, USA, 5-9 May 2014, 90711L, 2014.

Fang, L., He, N., Li, S., Plaza, A. J., and Plaza, J.: A New Spatial–Spectral Feature Extraction Method for Hyperspectral Images Using Local Covariance Matrix Representation, IEEE Trans. Geosci. Remote Sens., 56, 3534-3546, 10.1109/TGRS.2018.2801387, 2018.

Faraki, M., Harandi, M. T., and Porikli, F.: More about VLAD: A leap from Euclidean to Riemannian manifolds, in: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition, 7-12 June 2015, 4951-4960, 2015.

Faraki, M., Palhang, M., and Sanderson, C.: Log-Euclidean Bag of Words for Human Action Recognition, IET Comput. Vision, 9, 331-339, 10.1049/iet-cvi.2014.0018, 2015b.

Guo, K., Ishwar, P., and Konrad, J.: Action Recognition Using Sparse Representation on Covariance Manifolds of Optical Flow, in: Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, 29 Aug.-1 Sept. 2010, 188-195, 2010.

Hartmann, D. L., Ockert-Bell, M. E., and Michelsen, M. L.: The Effect of Cloud Type on Earth's Energy Balance: Global Analysis, J. Clim., 5, 1281-1304, 10.1175/1520-0442(1992)005<1281:Teocto>2.0.Co;2, 1992.

Heinle, A., Macke, A., and Srivastav, A.: Automatic cloud classification of whole sky images, Atmos. Meas. Tech. Discuss., 3, 557-567, 10.5194/amtd-3-269-2010, 2010.

Hsu, C.-W. and Lin, C.-J.: A Comparison of Methods for Multiclass Support Vector Machines, IEEE Trans. Neural Networks Learn. Syst., 13, 415-425, 10.1109/72.991427, 2002.

Huang, Z., Wang, R., Shan, S., Gool, L. V., and Chen, X.: Cross Euclidean-to-Riemannian Metric Learning with Application to Face Recognition from Video, IEEE Trans. Pattern Anal. Mach. Intell., 40, 2827-2840, 10.1109/TPAMI.2017.2776154, 2018.

J. Huertas, J. R.-B., D. Pozo, R. Aler, Inés M. Galván: Genetic programming to extract features from the whole-sky camera for cloud type classification, in: Proceedings of the International Conference on Renewable Energies and Power Quality, Malaga, Spain, 4-6 April 2017, 132-136, 2017.

Jégou, H., Douze, M., and Schmid, C.: Improving Bag-of-Features for Large Scale Image Search, Int. J. Comput. Vision, 87, 316-336, 10.1007/s11263-009-0285-2, 2010.

Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P., and Schmid, C.: Aggregating local image descriptors into compact codes, IEEE Trans. Pattern Anal. Mach. Intell., 34, 1704-1716, 10.1109/TPAMI.2011.235, 2012.

Joubert, A.: The astronomical image - Toward an objective analysis, Lastronomie, 93, 3-30, 1978.

Kazantzidis, A., Tzoumanikas, P., Bais, A., Fotopoulos, S., and Economou, G.: Cloud Detection and Classification with the Use of Whole-Sky Ground-Based Images, Atmos. Res., 113, 80-88, 10.1016/j.atmosres.2012.05.005, 2012.

Knerr, S., Personnaz, L., and Dreyfus, G.: Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: Neurocomputing： Algorithms, Architectures and Applications, Springer, Berlin, Heidelberg, Germany, 41-50, 1990.

Kreßel, U. H.-G.: Pairwise classification and support vector machines. In: Advances in kernel methods: support vector learning, MIT Press, Cambridge, 255–268, 1999.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, USA, 3-8 December 2012, 1097–1105, 2012.

Li, M., Liu, S., and Zhang, Z.: Deep tensor fusion network for multimodal ground-based cloud classification in weather station networks, Ad Hoc Networks, 96, 101991, 10.1016/j.adhoc.2019.101991, 2020.

Li, P., Wang, Q., Zuo, W., and Zhang, L.: Log-Euclidean Kernels for Sparse Representation and Dictionary Learning, in: Proceedings of 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1-8 Dec 2013, 1601-1608, 2013.

Liu, L., Sun, X., Chen, F., Zhao, S., and Gao, T.: Cloud Classification Based on Structure Features of Infrared Images, J. Atmos. Oceanic Technol., 28, 410-417, 10.1175/2010jtecha1385.1, 2011.

Liu, L., Sun, X., Gao, T., and Zhao, S.: Comparison of Cloud Properties from Ground-Based Infrared Cloud Measurement and Visual Observations, J. Atmos. Oceanic Technol., 30, 1171-1179, 10.1175/jtech-d-12-00157.1, 2013.

Liu, S., Duan, L., Zhang, Z., and Cao, X.: Hierarchical Multimodal Fusion for Ground-Based Cloud Classification in Weather Station Networks, IEEE Access, 7, 85688-85695, 10.1109/ACCESS.2019.2926092, 2019.

Liu, S. and Li, M.: Deep multimodal fusion for ground-based cloud classification in weather station networks, EURASIP J WIREL COMM, 2018, 48, 10.1186/s13638-018-1062-0, 2018.

Liu, S., Li, M., Zhang, Z., Xiao, B., and Cao, X.: Multimodal Ground-Based Cloud Classification Using Joint Fusion Convolutional Neural Network, Remote Sens., 10, 822, 10.3390/rs10060822, 2018.

Liu, S., Zhang, Z., and Mei, X.: Ground-based cloud classification using weighted local binary patterns, J. Appl. Remote Sens., 9, 095062, 10.1117/1.JRS.9.095062, 2015.

Luo, Q., Meng, Y., Liu, L., Zhao, X., and Zhou, Z.: Cloud classification of ground-based infrared images combining manifold and texture features, Atmos. Meas. Tech., 11, 5351-5361, 10.5194/amt-11-5351-2018, 2018.

Pang, Y., Yuan, Y., and Li, X.: Gabor-Based Region Covariance Matrices for Face Recognition, IEEE Trans. Circuits Syst. Video Technol., 18, 989-993, 10.1109/TCSVT.2008.924108, 2008.

Pennec, X.: Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements, J. Math. Imaging Vision, 25, 127-154, 10.1007/s10851-006-6228-4, 2006.

Pennec, X., Fillard, P., and Ayache, N.: A Riemannian Framework for Tensor Computing, Int. J. Comput. Vision, 66, 41-66, 10.1007/s11263-005-3222-z, 2006.

Ramanathan, V., Cess, R., Harrison, E. F., Minnis, P., Barkstrom, R. B., Ahmad, E., and Hartmann, D.: Cloud-radiative forcing and climate: Results from the Earth's radiation budget, Science, 243, 57-63, 10.1126/science.243.4887.57, 1989.

Shi, C., Wang, C., Wang, Y., and Xiao, B.: Deep Convolutional Activations-Based Features for Ground-Based Cloud Classification, IEEE Geosci. Remote Sens. Lett., 14, 816-820, 10.1109/LGRS.2017.2681658, 2017.

Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, in: Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7-9 May 2015, 1-14, 2015.

Sra, S.: A new metric on the manifold of kernel matrices with application to matrix geometric means, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, USA, 3-8 December 2012, 144–152, 2012.

Taravat, A., Frate, F. D., Cornaro, C., and Vergari, S.: Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images, IEEE Geosci. Remote Sens. Lett., 12, 666-670, 10.1109/LGRS.2014.2356616, 2015.

Tsai, C.-F.: Bag-of-Words Representation in Image Annotation: A Review, ISRN Artificial Intelligence, 2012, 376804, 10.5402/2012/376804, 2012.

Tuzel, O., Porikli, F., and Meer, P.: Region Covariance: A Fast Descriptor for Detection and Classification, in: Proceedings of the 9th European Conference on Computer Vision, Berlin, Heidelberg, 7-13 May 2006, 589-600, 2006.

Wang, R., Guo, H., Davis, L. S., and Dai, Q.: Covariance discriminative learning: A natural and efficient approach to image set classification, in: Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16-21 June 2012, 2496-2503, 2012.

Wang, W., Wang, R., Huang, Z., Shan, S., and Chen, X.: Discriminant Analysis on Riemannian Manifold of Gaussian Distributions for Face Recognition with Image Sets, IEEE Trans. Image Process., 27, 151-163, 10.1109/TIP.2017.2746993, 2018a.

Wang, Y., Shi, C., Wang, C., and Xiao, B.: Ground-based cloud classification by learning stable local binary patterns, Atmos. Res., 207, 74-89, 10.1016/j.atmosres.2018.02.023, 2018b.

Xiao, Y., Cao, Z., Zhuo, W., Ye, L., and Zhu, L.: mCLOUD: A Multi-view Visual Feature Extraction Mechanism for Ground-based Cloud Image Categorization, J. Atmos. Oceanic Technol., 33, 789, 10.1175/JTECH-D-15-0015.1, 2016.

Ye, L., Cao, Z., and Xiao, Y.: DeepCloud: Ground-Based Cloud Image Categorization Using Deep Convolutional Features, IEEE Trans. Geosci. Remote Sens., 55, 5729-5740, 10.1109/TGRS.2017.2712809, 2017.

Zhang, Z., Li, D., and Liu, S.: Salient Dual Activations Aggregation for Ground-Based Cloud Classification in Weather Station Networks, IEEE Access, 6, 59173-59181, 10.1109/ACCESS.2018.2874994, 2018.

Zhuo, W., Cao, Z.-G., and Xiao, Y.: Cloud Classification of Ground-Based Images Using Texture–Structure Features, J. Atmos. Oceanic Technol., 31, 79-92, 10.1175/JTECH-D-13-00048.1, 2014.